



Evaluation of 20CR reanalysis data and model results based on historical (1930–1940) observations from Franz Josef Land

Daniel KLAUS¹, Przemysław WYSZYŃSKI^{2,*}, Klaus DETHLOFF¹,
Rajmund PRZYBYŁAK² and Annette RINKE¹

¹ Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research,
Telegrafenberg A43, 14473 Potsdam, Germany

² Nicolaus Copernicus University, Department of Meteorology and Climatology,
Lwowska 1, 87-100 Toruń, Poland

* corresponding author <Przemyslaw.Wyszynski@umk.pl>

Abstract: Unique and independent historical observations, carried out in the central Arctic during the early twentieth century warming (ETCW) period, were used to evaluate the older (20CRv2) and newer (20CRv2c) versions of the 20th Century Reanalysis and the HIRHAM5 regional climate model. The latter can reduce several biases compared to its forcing data set (20CRv2) probably due to higher horizontal resolution and a more realistic cloud parameterization. However, low-level temperature and near-surface specific humidity agree best between 20CRv2c and the surface-based observations. This better performance results from more realistic lower boundary conditions for sea ice concentration and sea surface temperature, but it is limited mainly to polar night. Although sea level pressures are very similar, the vertical stratification and baroclinicity change in the transition from 20CRv2 to 20CRv2c. Compared to observed temperature profiles, the systematic cold bias above 400 hPa remains almost unchanged indicating an incorrect coupling between the planetary boundary layer and free troposphere. In addition to surface pressures, it is therefore recommended to assimilate available vertical profiles of temperature, humidity and wind speed. This might also reduce the large biases in 10 m wind speed, but the reliability of the sea ice data remains a great unknown.

Key words: Arctic, 20th Century Reanalysis, regional climate model, early twentieth century warming.

Introduction

The remarkable near-surface temperature increase in high latitudes from about 1920 to 1940 is commonly referred to as early twentieth century warming (ETCW). For instance, Scherhag (1937) compared the extended winter (from

November to March) temperature of ten-year periods to a long-term mean (1876–1925) and found that the air in Jakobshavn (Greenland) was 5°C warmer in the time slice 1923–1932 than in the time slice 1883–1892. Similarly, Scherhag (1937) quantified the ETCW for Green Harbour (Spitsbergen), where he calculated a temperature increase of 9°C between the time slices 1911–1915 and 1931–1935.

In their review article, Wood and Overland (2010) gave a comprehensive overview of this Arctic warming period that interestingly shows many similarities to the warming observed in more recent years (1981–2010). Both the knowledge of climate conditions prevailing during the ETCW and the identification and understanding of its major causes are of crucial importance for correctly interpreting the current warming in the Arctic. Consequently, there is renewed interest in the insufficiently understood mechanisms leading to the ETCW (Wood and Overland 2010; Semenov and Latif 2012).

Initially, it was broadly accepted that the ETCW arose mainly due to changes in the large-scale atmospheric circulation (Scherhag 1937; Zubov 1948; Przybylak 2016), but modeling studies brought forward new ideas. While Scott *et al.* (2003) as well as Hegerl *et al.* (2003) concluded that the ETCW is attributable to an increase of well-mixed greenhouse gases, the authors identified changes in solar irradiance and volcanic activity, respectively, as additional factor. The modeling studies of Nozawa *et al.* (2005) and Shiogama *et al.* (2006) were indicative that natural factors, *i.e.* changes in solar irradiance and volcanic influences, are of greatest importance. The latter result was basically confirmed by Suo *et al.* (2013) who analyzed model simulations and observations. On the other hand, Suo *et al.* (2013) argued that the internal variability of the Arctic climate system was less important for the ETCW, while Meehl *et al.* (2003) found that it played a major role.

The impact of Arctic sea ice changes in the past is currently under controversial discussion. The temporal evolution of atmospheric circulation anomalies following Arctic sea ice loss in autumn starts with a strengthening and north-westward shift of the Scandinavian High at the surface and the mid-troposphere and the Siberian High at the surface in early winter. This wave train pattern leads to enhanced vertical planetary wave fluxes for low ice conditions as discussed by Jaiser *et al.* (2016). It reduces the strength of the polar vortex in the stratosphere and leads to a negative Arctic Oscillation/North Atlantic Oscillation circulation response in the troposphere, either by planetary wave breaking and wave reflection or by potential vorticity inversion with downward propagation within weeks and adjustment of the circulation as shown in model sensitivity studies by Nakamura *et al.* (2015). These complex tropo-stratospheric linkages are not yet understood and their impact on historical climate changes like the ETCW was never discussed. Even today there is a controversial discussion on the major causes leading to the ETCW (Nozawa *et al.* 2005), and it is still not entirely clear whether they originated from natural or anthropogenic

sources. Nevertheless, the vast majority of related studies give evidence that natural factors played the dominant role (Bengtsson *et al.* 2004; Nozawa *et al.* 2005; Shiogama *et al.* 2006; Wood and Overland 2010; Suo *et al.* 2013). The current knowledge about ETCW is appropriately summarized by the words of Bengtsson *et al.* (2004): “The 1920–1940 Arctic warming is one of the most puzzling climate anomalies of the 20th century.”

Further progress and a satisfying compromise regarding the main drivers of ETCW are impeded by the sparsely available meteorological observations and limited information on the hydrosphere and cryosphere (*e.g.*, Bengtsson *et al.* 2004). Wood and Overland (2010) pointed to the fact that all these observational data either contain gaps or have insufficient temporal or spatial resolution. Thus, the crucial problem of gridded data sets for sea ice concentration and sea surface temperature (SST), *e.g.* HadISST1.1 (Rayner *et al.* 2003) from the United Kingdom Met Office Hadley Centre, is the use of a limited number of observations and interpolation techniques (Bengtsson *et al.* 2004). In this regard, Wood and Overland (2010) hypothesized the unreliability of available surface forcing data used as lower boundary conditions (BCs) in a climate model, and they mentioned rather negligible changes in the sea ice data from HadISST1.1 prior to 1960. However, Przybylak (2002), Bengtsson *et al.* (2004), and Wood and Overland (2010) identified a rather small but well recognized negative correlation (around -0.6) between surface air temperature and sea ice concentration during the ETCW. Zakharov (2003) and Johannessen *et al.* (2004) demonstrated the significant decline in sea ice at that times for the whole Arctic and Alekseev *et al.* (2009) for the eastern Arctic and Nordic Seas. Recently, Walsh *et al.* (2016) found the same negative trend for the whole Arctic, but they also pointed out that the detected loss of sea ice during the ETCW depends on the marginal sea of the Arctic Ocean.

A new and complementary effort to analyze the climate during the 20th century is the generation of global reanalysis products that are based on the assimilation of historical observations for a few quantities, *e.g.* surface pressure, and the prescription of SSTs and sea ice concentrations. Prominent examples are ERA-20C (Poli *et al.* 2013) from the European Centre for Medium-Range Weather Forecasts and the “20th Century Reanalysis” (20CR; Compo *et al.* 2011) generated by National Oceanic and Atmospheric Administration and the University of Colorado Cooperative Institute for Research in Environmental Sciences (NOAA-CIRES). Since only the latter is analyzed in the present study it is explained in detail in the section *Reanalysis data*.

The performance of 20CR in the Arctic has been the subject of several studies (Brönnimann *et al.* 2013; Krueger *et al.* 2013; Przybylak *et al.* 2013, 2016). In this regard, the developer of 20CR (Compo *et al.* 2011) already provided the ensemble uncertainty of sea level pressure (SLP) in the Arctic area, which is tenfold larger in historical times, when data coverage is poorer, than in the

present period; for details see their fig. 1. Przybylak *et al.* (2013) revealed quite a large positive bias of the *SLP* at about 4 hPa in the Arctic in the period 1872–1920 in comparison to the real data from the instrumental observations (see their table VIII). Based on air temperature measurements in the sea around Svalbard, Przybylak *et al.* (2016) found that reanalyses (ERA-20C and 20CR) are usually too cold in comparison with observations, except some southern grid boxes in the case of 20CR. In particular, large differences (2–5°C) have been found in northern regions; see their fig. 9. On the other hand, Brönnimann *et al.* (2013) and Krueger *et al.* (2013) compared reanalyzed horizontal wind speeds to observations and the latter authors even analyzed storminess. However, any explicit assessment of the reanalyzed vertical temperature structure, the stratification in the planetary boundary layer (PBL), the reproducibility of baroclinicity, and the coupling between PBL and free troposphere during the ETCW is missing until now. It is particularly questionable whether 20CR can produce the linkage between baroclinicity and the large-scale circulation proposed by Jaiser *et al.* (2012) keeping in mind the assimilation of limited surface observations. Otherwise, there is need of initial analyzes regarding the role of Arctic clouds for the climate conditions during the ETCW.

For the older and newer version of 20CR, only surface pressure observations were assimilated and all other fields were computed based on the underlying atmospheric model equations. Therefore, an independent check of the quality of these atmospheric variables by using data sets based directly on observations like radiosonde observations are of eminent importance. Recently, Poli *et al.* (2016) introduced a new atmospheric reanalysis product of the twentieth century, the so-called ERA-20C that covers the period from 1900–2010. It was generated by a different model with finer horizontal and vertical resolution, where not only surface pressure observations but also additional wind observations over the oceans were assimilated improving the realism of synoptic-scale processes (Poli *et al.* 2016). Poli *et al.* (2016) also compared ERA-20C against 20CRv2c and found that the variance of both data sets is comparable even though the lack of more detailed information in the poorly observed Arctic.

Before it is hopefully possible to correctly identify the causes of ETCW from historical reconstructions, it is mandatory to uncover their strengths and weaknesses in a similar way as in case of climate models. The present study assesses the quality of reanalyzed and modeled (near-)surface quantities like 2 m air temperature, 2 m specific humidity, 10 m horizontal wind speed, and addresses all issues mentioned in the previous paragraph. In particular, surface-based historical observations including tracked balloon and radiosonde ascents, carried out at Calm Bay on Franz Josef Land (Fig. 1), were exploited to evaluate two versions of 20CR and a regional climate model (RCM) simulation for the ETCW period.

Data description

Surface observations. — Meta data of historical surface observations used in the present study were obtained from the meteorological report (Lvov 1933) on the overwintering 1930–1931 in an Arctic measuring station on Franz Josef Land (Fig. 1), established by the Arctic and Antarctic Research Institute (AARI; “Arctic Institute of the USSR” prior to 1958). In particular, this meteorological station was located in the Calm Bay (80.32°N, 52.80°E) on Hooker Island. The latter is one of the comparatively large islands of the archipelago and almost entirely covered with an ice cap. Due to its location north of the Calm Bay and the British Channel, it is protected from the northerly and northeasterly winds by nearly vertical slopes of coastal rocks of about 140 m height. The site of the exposure of instruments was situated on Cape Sedoff.

The following measuring instruments were utilized to determine the atmospheric air pressure: (1) a cistern barometer kept in a special casing fixed inside on the main wall of the station building; (2) a station aneroid placed on the lower shelf of the barometer casing; and (3) two barographs, a diurnally and a weekly one, installed on the special shelves and fixed to the same wall as the

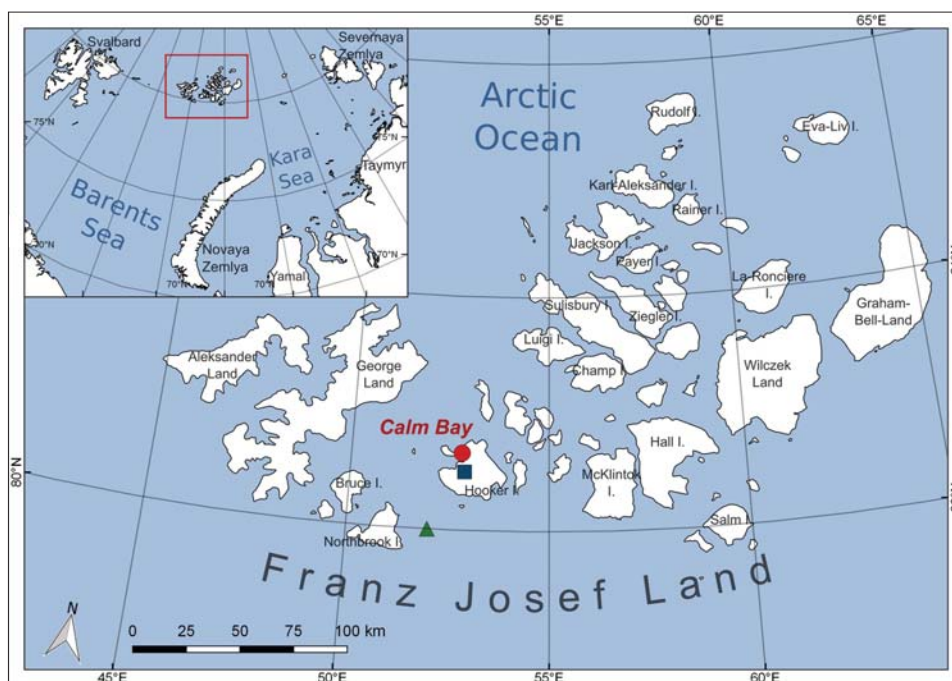


Fig. 1. Location of the historical meteorological station Calm Bay (dot) on Hooker Island belonging to Franz Josef Land as well as the nearest-neighbor grid points of the 20th Century Reanalysis (triangle) and regional climate model HIRHAM5 (square), respectively.

barometer. The altitude of these instruments was approximately 5.9 m above sea level (a.s.l.). During the time period from 1 August to 25 December 1930, hourly observations were collected from the diurnal barograph. Due to the damage of its clockwork, the subsequent (beginning of 25 December to 31 July 1931) hourly pressure measurements needed to be read from the weekly barograph.

Two meteorological screens with louvered construction were installed 2.3 m above the ground. The first screen contained a psychrometer and the second various self-recording instruments, *i.e.* thermographs and hygrographs. This enabled measurements of water vapor pressure (e), which were carried out three times per day, at 7, 13, and 21 LMT (local mean time).

Normally, air temperature with hourly resolution was obtained according to the readings of the diurnal thermograph. From 24 March to 29 May 1931, temperature measurements were taken from the weekly thermograph, because the clockwork of the diurnal thermograph was in repair. The weekly thermograph worked continuously throughout the year.

A wind vane with a heavy board was installed on a wooden column in 9 m height to estimate wind direction, while the horizontal wind speed (V) was measured three times per day, at 7, 13 and 21 LMT. The total cloud cover ($TCLC$) was visually observed with the same chronology and in the historically well-known tenth scale, where 0/10 (= 0%) means clear sky and 10/10 (= 100%) overcast.

Processing the historical surface observations. — For avoiding errors, the observational data (Table 1) were extracted manually into digital format. This kind of approach guarantees the lowest error rates when acquiring data from antique books and prints (Brönnimann *et al.* 2006). All outliers were checked and suspect ones removed after inspection from the data set. However, it is worth noting here that the first quality control procedure was performed by the meteorological observer A. J. Goloubenkoff already during the expedition. A further quality control was performed by the staff of the section for Polar and High Altitude Observations of the Central Geophysical Observatory in St. Petersburg.

Barographs and aneroids are compensated for temperature changes and they are also not affected by gravity effects. Thus, a gravity correction needed not to be applied to the actual pressure reading. The sea level pressure (SLP) was calculated according to appendix 2 of Cappelen (2009):

$$SLP = p \left(1 + \frac{g}{R_d} \times \frac{h_B}{(T/10) + T_{melt}} \right), \quad (1)$$

where p represents the air pressure at station level (accuracy of 0.1 hPa), h_B the height of the barograph in meters a.s.l., and T the air temperature at station level (accuracy of 0.1 K). Furthermore, $g = 9.82 \text{ ms}^{-2}$ is the gravitational acceleration, $R_d = 287.04 \text{ J(kgK)}^{-1}$ the specific gas constant of dry air, and $T_{melt} = 273.15 \text{ K}$ the melting temperature of fresh water.

Table 1

Geographic location of the historical station Calm Bay and the nearest-neighbor grid points of the 20th Century Reanalysis (versions 20CRv2 and 20CRv2c) and the HIRHAM5 model. Temporal usage and resolution of raw data for sea level pressure (SLP), horizontal wind speed (V_{10m}), air temperature (T), specific humidity (SH_{2m}) and total cloud cover ($TCLC$).

Station/ Nearest grid point	Longitude	Latitude	Temporal data usage	SLP (hPa)	V_{10m} (ms^{-1})	T ($^{\circ}C$)	SH_{2m} (gkg^{-1})	$TCLC$ (%)
Calm Bay (<i>surface observations</i>)	52.80°E	80.32°N	1930.08.01 – 1931.07.31	h ¹	t ²	h ³	t	t
Calm Bay (<i>vertical profiles</i>)			1934.09.10 – 1940.12.31	–	–	i ⁴	–	–
20CRv2c	52.00°E	80.00°N	1915.01.01 – 1940.12.31	6-h ⁵	6-h	6-h	6-h	6-h
20CRv2				6-h	6-h	6-h	6-h	6-h
HIRHAM5	52.88°E	80.25°N		6-h	6-h	6-h	6-h	6-h

¹ hourly, ² three times a day (7, 13, 21 LMT) in 10 m a.g.l., ³ hourly in 2 m a.g.l., ⁴ irregular from 1000 hPa to 10 hPa, ⁵ 6-hourly at the corresponding height of observational data

Specific humidity (SH_{2m}) in 2 m above ground level (a.g.l.) was recalculated based on equation (8) in chapter III by Rózdzyński (1999):

$$SH_{2m} = 1000\varepsilon \times e/p = 621.98 \times e/p, \quad (2)$$

where e represents the water vapor pressure measured by psychrometer. The dimensionless constant $\varepsilon = 0.62198$ is the ratio of the relative molar masses of water and dry air and it is multiplied by 1000 to obtain the conventional unit (gkg^{-1}).

The correction for obtaining the horizontal wind speed at 10 m a.g.l. (V_{10m}) was introduced according to the Hellmann's formula used in equation (63) of chapter IV by Rózdzyński (1999):

$$V_{10m} = V/[0.233 + 0.656 \times \lg(h_A + 4.75)], \quad (3)$$

where h_A is the height of the anemometer in meters a.g.l. and V the actually measured wind speed.

Daily means (DMs) of SLP and T_{2m} were calculated by use of

$$DM = (Q_{t1} + Q_{t2} + \dots + Q_{t24})/24, \quad (4)$$

while in case of V_{10m} , SH_{2m} and $TCLC$ the following relation

$$DM = (Q_{t7} + Q_{t13} + Q_{t21})/3, \quad (5)$$

was used due to measurements only three times a day. In the equations (4) and (5) Q_{t1} , Q_{t2} , ..., Q_{t24} are respective values of a quantity at 1, 2, ..., 24 LMT.

Vertical profiles of air temperature. — Observations of the vertical temperature structure at Calm Bay were taken from PANGAEA – Data Publisher for Earth & Environmental Science. This digital data library provides a historical radiosondes and tracked balloons archive on standard pressure levels back to the 1920s (Ramella Pralungo *et al.* 2014). In particular, files are available in NetCDF format including observed temperatures interpolated to 16 pressure levels: (1000, 925, 850, 700, 500, 400, 300, 250, 200, 150, 100, 70, 50, 30, 20, 10) hPa.

In case of Calm Bay, PANGAEA provides 12-hourly vertical profiles from 10 September 1934 to 21 April 1955. Particularly due to a limited number of expeditions, this data set contains a large number of missing temperature profiles during the first 5–10 years. The observations were usually carried out daily but sometimes only every other day, resulting in additional data gaps. In fact, an observed vertical temperature profile can be available actually twice (0 and 12 UTC) at a certain day but also either at 0 UTC or at 12 UTC. Even if a vertical temperature profile was actually measured a missing value can appear either in a single or multiple pressure levels. Comparing these inhomogeneous and irregular observations on a 12-hourly time base to reanalyzed and simulated temperature profiles would be very time consuming. Thus, multi-year monthly mean profiles of air temperature were derived from the original PANGAEA data also to allow more general statements. Here, the calculation of daily mean values was followed by the calculation of multi-year means. According to Table 1, their calculation was based on temperature profiles measured from September 1934 to December 1940. This guaranteed an overlapping time period with the reanalysis and model data, which are explained in the next two sections.

Reanalysis data. — The “Twentieth-Century Reanalysis, Version 2” (20CRv2; Compo *et al.* 2011), generated by the Physical Sciences Division (PSD) of the Earth System Research Laboratory (ESRL) from National Oceanic and Atmospheric Administration (NOAA) and the Cooperative Institute for Research in Environmental Sciences (CIRES) at the University of Colorado, is a historical and comprehensive global atmospheric circulation data set that covers the time period from 1871–2012. Its development was mainly motivated by the need for providing an observational validation data set with quantified uncertainties. This should facilitate the assessment of climate model simulations for the 20th century

with focus on the statistics of daily weather (NCAR 2017). Regarding 20CR our analyses used mainly the mean fields and the variance fields to a smaller extent each based on the 56 member ensemble (details given by Compo *et al.* 2011). These fields are archived routinely by ESRL/PSD of NOAA-CIRES (https://www.esrl.noaa.gov/psd/data/20thC_Rean/).

The assimilation procedure of 20CRv2 used available measurements of surface pressure or *SLP* and in addition monthly Hadley Centre (HadISST1.1) sea ice concentrations (c_{ice}) and SSTs as lower BCs. In particular, an Ensemble Kalman Filter data assimilation method was applied, while the National Centers for Environmental Prediction (NCEP) global numerical weather prediction land/atmosphere model contributed background fields as “first guess”. This NCEP model has a Gaussian grid with horizontal resolution of $\sim 2^\circ$ latitude \times 1.875° longitude (T62 with 94×192 grid points), and it applies 28 model levels with top at 10 hPa.

More recently, an upgraded version (20CRv2c) was generated by use of the same model but with improved lower BCs and a temporal extension back to 1850. While c_{ice} was previously truncated at 0.5, *i.e.* 50%, the NCEP model permits the specification down to fractions of 0.15 in 20CRv2c. Additionally, the misspecification of c_{ice} noted by Compo *et al.* (2011) is corrected. The improved lower BCs comprise monthly fields of COBE-SST2 sea ice concentrations (Hirahara *et al.* 2014) and SSTs from an ensemble of pentad Simple Ocean Data Assimilation with Sparse Input version 2 (Giese *et al.* 2016) corrected to COBE-SST2 for latitudes $>60^\circ\text{N}$ (SODAsi.2c). Further, additional surface pressure observations, contained in the International Surface Pressure Databank version 3 (Compo *et al.* 2015), were taken into account. So far, the surface pressures mentioned in Section Surface observations were not assimilated.

While c_{ice} is prescribed in the NCEP Global Forecast System, sea ice thickness (d_{ice}) and temperature are derived from a thermodynamic three-layer sea ice model that was reformulated by Winton (2000). Further, the NCEP model applies a relative humidity based cloud scheme for the diagnostic calculation of fractional cloud cover.

Both in 20CRv2 and 20CRv2c, the provided analysis (*e.g.*, *SLP*) and forecast (*e.g.*, V_{10m} , T_{2m} , SH_{2m} , $TCLC$) fields, partly listed in Table 1, have different spatial coverage. In particular, the analysis fields are available on a regular $2^\circ \times 2^\circ$ global grid from 90°N to 90°S latitude and from 0°E to 360°E longitude (91×180 grid points). In contrast, the forecast fields are provided on the Gaussian grid covering 88.542°N – 88.542°S and 0°E – 358.125°E , respectively. 20CR data were extracted at the nearest neighbor grid point to Calm Bay (see Table 1 and Fig. 1), which is actually an ocean point. In fact, Franz Josef Land doesn’t exist in the land sea mask of 20CR. The analysis (forecast) fields were generated 6-hourly (3-hourly) and were the basis for the calculation of daily and monthly means.

Model data. — The atmospheric RCM “HIRHAM5” combines the dynamical core of the regional weather forecast model HIRLAM7 (High Resolution Limited Area Model; Undén *et al.* 2002) and the physical parameterizations of the atmospheric general circulation model ECHAM5 (European Centre Hamburg model; Roeckner *et al.* 2003). It has a pan-Arctic domain that includes most areas north of 60°N. The model uses an Eulerian-type time discretization with a time step of 2 min, a horizontal resolution of 0.25° (~25 km), and 40 vertical levels in σ -p-coordinates (Berrisford *et al.* 2009) that cover an altitude range from about 10 m above the surface up to 10 hPa. The lowermost 1 km is represented by 10 levels.

HIRHAM5 was initialized and run with 6-hourly 20CRv2 forcing at the lateral boundaries (Davies 1976). In addition, a nudging technique was applied, except for the boundary zone, to adopt the large-scale circulation of 20CRv2. The latter can be assumed as realistic and reasonable representation of the historic state even though considering the larger uncertainty in data sparse regions discussed by Compo *et al.* (2011). In fact, the “dynamical relaxation” technique (Davies and Turner 1977) is used, where the nudging strength increases linearly with altitude (from 0% in the lowest to 1% in the topmost model level). The nudging strength means the percentage of the 20CRv2 value to the overall model solution at a certain grid point, level and time step.

SSTs and c_{ice} were prescribed based on 20CRv2 fields, where a constant value of 2 m was assumed for the sea ice thickness. This means that the ice-covered part of a certain grid cell has 2 m thick sea ice. In contrast to Roeckner *et al.* (2003), HIRHAM5 uses the first version of the sea ice albedo scheme of Køltzow (2007) that takes into account the effects of melt ponds during summer and snow on sea ice in winter/spring. The standard prognostic-statistical cloud scheme as explained by Klaus *et al.* (2016) is applied for diagnostically calculating fractional cloud cover. Klaus *et al.* (2016) also provide more detailed information about the physical parameterizations and further differences to Roeckner *et al.* (2003).

A 26 years (from 1 January 1915 to 31 December 1940) model simulation was conducted that includes the ETCW period. For the comparison to reanalysis and observational data, the model results were extracted at the nearest neighbor grid point to Calm Bay, which is actually a land point (see Table 1 and Fig. 1). Analogous to the reanalysis data, the 6-hourly model results were used to calculate daily and monthly means.

Results

Before the comparison of 20CR (both v2 and v2c) to independent historical data, their different performance during historical and contemporary time periods should be mentioned for the site of Calm Bay on Franz Josef Land. Based on the

spread, meant as a ensemble standard deviation, of near-surface meteorological variables provided by ESRL/PSD of NOAA-CIRES (see the section *Reanalysis data*) monthly and daily long term mean ensemble spreads have been calculated for the periods 1911–1940 and 1981–2010. Mean spreads in 20CRv2 and 20CRv2c, both daily and monthly, are comparable. The performance of 20CR in present times is from 3 up to even 10 times better than in the historical period when a significantly lesser amount of observational data has been assimilated. For instance, the mean monthly ensemble standard deviation of SLP , T_{2m} and $TCLC$ in 1981–2010 is 0.5–0.6 hPa, 0.5–0.7°C and 7%, respectively, whereas in the period 1911–1940 is 6–7 hPa, 4–6°C, and 21%, respectively.

Daily comparison to surface observations. — As claimed by NCAR (2017), 20CR was generated to evaluate climate models with focus on the statistics of daily weather. Figure 2 therefore illustrates the comparison of the observed, reanalyzed (20CRv2, 20CRv2c), and simulated (HIRHAM5) daily mean (A) SLP , (B) V_{10m} , (C) T_{2m} , (D) SH_{2m} , and (E) $TCLC$ for the overwintering 1930–1931 in Calm Bay. To allow more generalized statements, Fig. 2 also presents the results as 11-day moving average.

The observational data show an extremely variable SLP that varies between 980–1040 hPa. While wind speeds are usually small to moderate, significantly larger values appear during the storm season from November to February when they often exceed 10 ms^{-1} . The observations show the expected annual cycle in near-surface air temperature that is driven by solar irradiation accompanied by colder temperatures in winter and warmer in summer. Near-surface specific humidity basically follows the observed curve of T_{2m} with the tendency that the warmer the temperature the moister is the air. $TCLC$ is in turn related to humidity, where the observations show larger cloud amounts in summer than in winter.

Cyclone activity can also be derived from the observational data in Fig. 2. At the beginning of January 1931, low $SLPs$ indicate a cyclone passage that coincides with larger wind speeds, warmer temperatures, moister air, and enhanced cloudiness. In contrast, the higher $SLPs$ in the second half of January 1931 indicate the presence of an anticyclone that coincides with smaller wind speeds, colder temperatures, dryer air, and less cloudiness.

There is a general agreement between observed, reanalyzed, and simulated $SLPs$ (Fig. 2A) but 20CRv2c produces the largest biases, especially from mid December to the end of April. At first glance, it is a little unexpected since the same surface pressures were assimilated as for the generation of the older reanalysis version (20CRv2). Due to the model forcing, $SLPs$ produced by HIRHAM5 and 20CRv2 are very similar. On the other hand, Figs. 3A to 3C show slightly higher correlation of daily $SLPs$ between the observations and 20CRv2c.

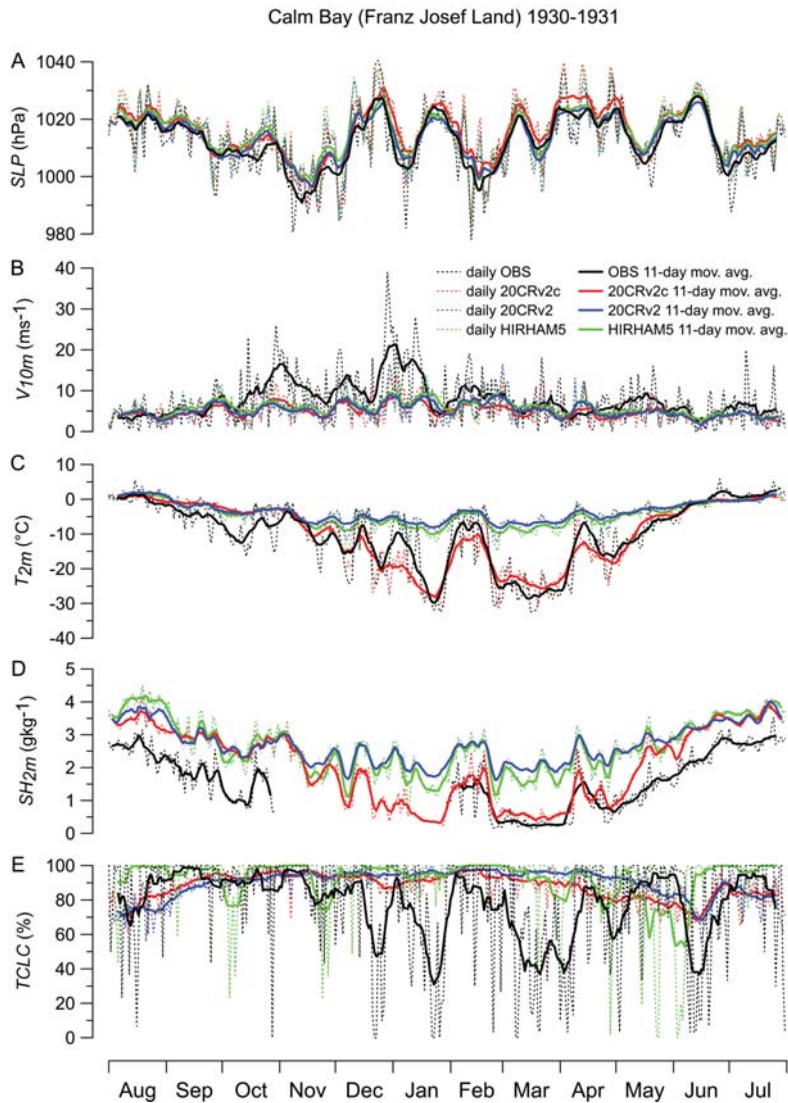


Fig. 2. Annual courses of the (A) sea level pressure; (B) 10 m horizontal wind speed; (C) 2 m air temperature; (D) 2 m specific humidity; and (E) total cloud cover from the surface-based observations (black), the two versions of 20CR (20CRv2 and 20CRv2c, blue and red, respectively), and the HIRHAM5 model (green) for Calm Bay from 1 August 1930 to 31 July 1931. Dotted lines indicate daily means, while thick lines represent the 11-day moving averages.

Compared to the measurements, V_{10m} is systematically underestimated (partly $>20 \text{ ms}^{-1}$ at the end of December 1930) by the two reanalysis versions and the model, especially from October 1930 to January 1931 (Fig. 2B). Additionally, Figs. 3D to 3F reveal comparatively small correlations to the observations ranging from 0.47 (HIRHAM5) to 0.56 (20CRv2c).

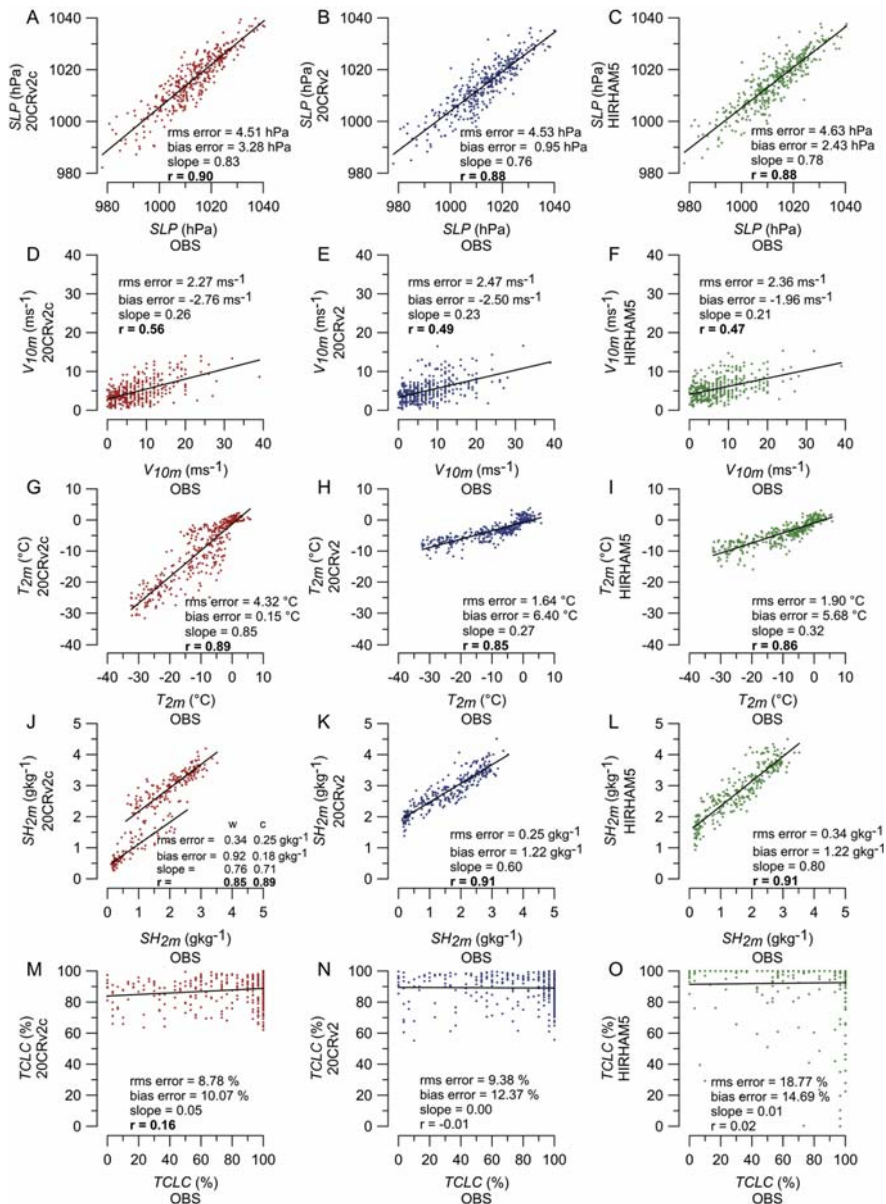


Fig. 3. Scatter plots of the (A–C) sea level pressure; (D–F) 10 m horizontal wind speed; (G–I) 2 m air temperature; (J–L) 2 m specific humidity; and (M–O) total cloud cover relating the surface-based observations with (A, D, G, J, M) the newer version of 20CR; (B, E, H, K, N) its older version; (C, F, I, L, O) the model for Calm Bay, based on daily data from 1 August 1930 to 31 July 1931. Beside the slope all subfigures include values for the root mean square (rms) error and bias error and correlation coefficients (r) in bold indicate significance on the $p \leq 0.05$ level. J is shown with separate linear regression lines and statistical measures: “w” – for the warm period (August 1930 – October 1930 and May 1931 – July 1931; top line) and “c” – for the cold period (February 1931 – April 1931; bottom line).

While 20CRv2 and HIRHAM5 show a reasonable annual cycle in T_{2m} (Fig. 2C), 20CRv2c can even produce a realistic annual amplitude of about 15°C . As a result, the newer version of 20CR shows a much smaller positive temperature bias to the observations. Interestingly, this clearly improved performance is only visible from November 1930 to May 1931. Exactly during this time period, HIRHAM5 is able to slightly reduce (up to 2 K) the positive T_{2m} bias compared to its forcing data set (20CRv2). Although the almost identical and high correlation values (Figs. 3G to 3I) argue for the qualitative agreement to the observations independent of the comparative data, the slope of the regression line is significantly closer to one in case of 20CRv2c.

Also independent of the data set, the annual cycle of SH_{2m} is in qualitative agreement to the observations (Fig. 2D). Correlations are generally large as well (Figs. 3J to 3L). The comparison of Figs 2C to 2D reveals that the curve shape of SH_{2m} is in all cases very similar to T_{2m} . Compared to the observations, the air in 2 m altitude is generally too moist, but to a much larger extent in case of HIRHAM5 and 20CRv2. HIRHAM5 is drier or slightly moister than 20CRv2 during polar night or polar day, respectively. In fact, 20CRv2c reproduces the absolute values of SH_{2m} fairly well during polar night, at least from February to April 1931, but the significant overestimation remains almost unchanged during the rest of the year.

Figure 2E clearly demonstrates that beside the predominant overestimation of $TCLC$ neither the two 20CR versions nor the model can capture the daily observations. The latter is confirmed by Figs. 3M to 3O, which show correlations <0.2 . Even though there is a mismatch to daily visual cloud observations, only HIRHAM5 is able to produce a noteworthy day-to-day variability of $TCLC$.

Monthly comparison to surface observations. — More meaningful on a climatological time scale is the comparison of observed monthly mean (A) SLP , (B) V_{10m} , (C) T_{2m} , (D) SH_{2m} , and (E) $TCLC$ to values produced by 20CR (20CRv2, 20CRv2c) and HIRHAM5, respectively. In addition to monthly averages (thick solid lines) for the overwintering 1930–1931 at Calm Bay, Fig. 4 illustrates the 2σ range (thin dotted lines) as an estimator for the variance.

These results underpin the findings obtained from the daily comparison. Figure 4A shows comparatively small differences to observations but larger SLP biases for 20CRv2c than 20CRv2, where the latter is very similar to HIRHAM5. Figure 4B shows a significant underestimation of V_{10m} and its variability independent of the comparative data set. Figure 4C shows the best reproduction of T_{2m} by 20CRv2c, but the variability is also too weak. Figure 4D shows that 20CRv2c cannot reduce the overestimation of SH_{2m} satisfactorily during polar day and the variability is still underrepresented. And finally, Fig. 4E shows a biased and too weak annual cycle of $TCLC$ for 20CR and HIRHAM5. However, selected month-to-month changes are more realistic in HIRHAM5, *e.g.*

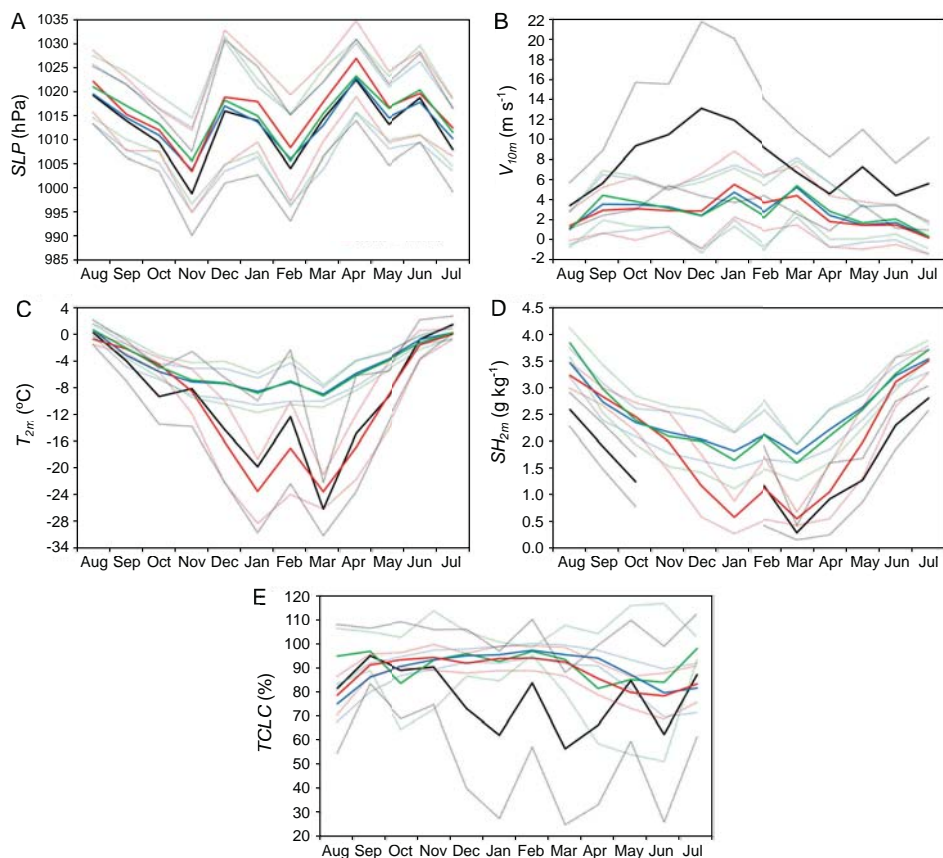


Fig. 4. Monthly averages of the (A) sea level pressure; (B) 10 m horizontal wind speed; (C) 2 m air temperature; (D) 2 m specific humidity; and (E) total cloud cover from surface-based observations (OBS, black), the newer version of 20CR (red), the older version (blue), and the model (green) for Calm Bay from August 1930 to July 1931. The thin lines indicate the corresponding 2σ range as a measure for the variability obtained from daily values.

for Sep-Oct and Jun-Jul compare black and green thick solid lines on Fig. 4E, and 20CR produces similar biases but without noteworthy variability.

For better quantification of the above-mentioned biases, Fig. 5 explicitly illustrates monthly differences of SLP , V_{10m} , T_{2m} , SH_{2m} , and $TCLC$ between 20CR (20CRv2, 20CRv2c) or HIRHAM5 and observations. These differences were calculated based on monthly means shown in Fig. 4.

Regarding SLP , the older version of 20CR produces the smallest bias to observations, and the model performance lies between them. While 20CRv2 shows predominantly differences <2 hPa, the bias of 20CRv2c is often >3 hPa, especially from January to May (Fig. 5A). In all months, both versions of 20CR and HIRHAM5 show a negative wind bias in the same order of magnitude

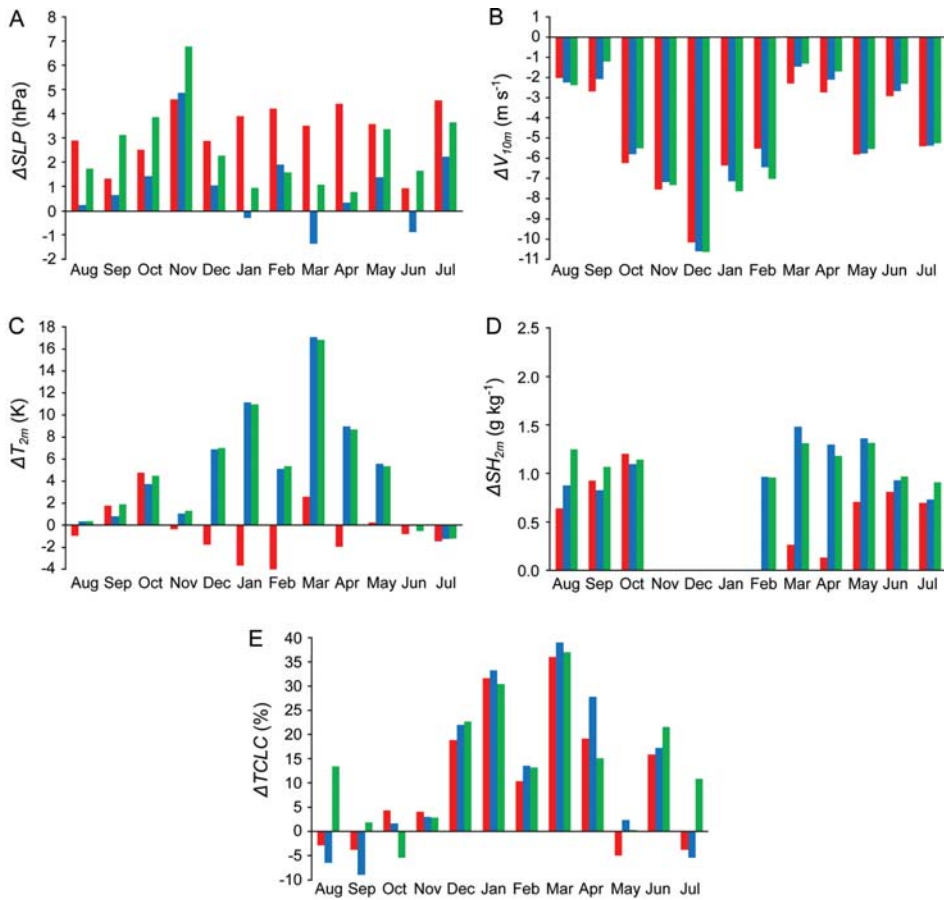


Fig. 5. Monthly mean (from August 1930 to July 1931) differences of the (A) sea level pressure; (B) 10 m horizontal wind speed; (C) 2 m air temperature; (D) 2 m specific humidity; and (E) total cloud cover based on Fig. 4. These differences were obtained by subtracting the observations at Calm Bay from values produced by a version of 20CR (20CRv2 in blue, 20CRv2c in red) or the model (HIRHAM5 in green).

(Fig. 5B). The largest differences to the observed V_{10m} appear from October to February with a maximum bias in December (around -10.5 ms^{-1}). Figure 5C shows a systematic and partly strong, around 17 K in March, positive bias in T_{2m} for 20CRv2 and HIRHAM5. This warm bias is partly reduced by the model. 20CRv2c produces smaller differences than HIRHAM5 and 20CRv2 only from November to May and also mostly a cold bias. SH_{2m} is systematically overestimated independent of whether reanalysis or model data (Fig. 5D). In summer, differences are between $0.5\text{--}1.0 \text{ g kg}^{-1}$ with slightly smaller values in case of 20CRv2c. The large overestimation of SH_{2m} ($>0.9 \text{ g kg}^{-1}$) in 20CRv2 and HIRHAM5 from February to May is significantly reduced in 20CRv2c

(differences $<0.25 \text{ gkg}^{-1}$). The improved performance of 20CRv2c regarding SH_{2m} is obviously related to reduced temperature biases. Thus, Figs. 5C and 5D also indicate a deterioration of SH_{2m} in October that coincides with an increased bias in T_{2m} .

Figure 5E shows an overestimation of $TCLC$ for HIRHAM5 and an underestimation of $TCLC$ for 20CR (20CRv2 and 20CRv2c) from July to September but an opposite behavior in October. Independent of the comparative data, the most significant overestimation of $TCLC$ (10–40%) occurs from December to April, when the bias in T_{2m} is large as well (Fig. 5C). A comparison of monthly differences in Fig. 5E according to amount reveals only a single month (October) with the smallest cloud bias for 20CRv2, while HIRHAM5 shows the smallest cloud bias in 5 months (September, November, January, April, May) and 20CRv2c in 6 months (August, December, February, March, June, July). However, the performance in terms of $TCLC$ is rather poor, independently of whatever reanalysis or model, at least for the analyzed overwintering 1930–1931.

Reanalyzed and simulated temperature profiles during the overwintering 1930–1931. — As expected, the daily and monthly comparison to surface-based observations confirm the best reproduction of T_{2m} by 20CRv2c. Nevertheless, this newer version of 20CR has several shortcomings, *e.g.* larger biases in SLP than its older version (20CRv2) or the model (HIRHAM5). The more realistic c_{ice} and d_{ice} , used as lower BCs for generating 20CRv2c, obviously improve T_{2m} but also change necessarily the vertical temperature structure and baroclinicity (*e.g.*, Jaiser *et al.* 2012).

Due to the unavailability of observed temperature profiles from August 1930 to July 1931 at Calm Bay, Fig. 6 illustrates the temporal evolution of daily averaged temperatures on 24 pressure levels extracted from 20CRv2, HIRHAM5, and 20CRv2c data. Although HIRHAM5 and its forcing data set (20CRv2) show almost the same temporal evolution of temperature (Figs. 6A and 6B), significant differences appear below 700 hPa (Fig. 6D). Temperature differences are also apparent above 700 hPa but have much smaller amount, except for the uppermost model levels. The latter is caused most likely by the missing gravity wave drag parameterization in the model that is switched off due to technical reasons.

The additional comparison to the newer version of 20CR (Fig. 6C) shows at least a similar temporal evolution of temperature, but results shown in Figs. 6E to 6F reveal much larger differences. Below about 950 hPa, 20CRv2c is significantly and systematically colder than 20CRv2 and HIRHAM5 (up to 23.6 K or 23.9 K in 1000 hPa at 27 January 1931) from December 1930 to May 1931. Although this basically agrees with the daily T_{2m} shown in Fig. 2C, there is no such clear systematics above 950 hPa. In contrast, differences between two versions of 20CR (Figs. 6E to 6F) indicate a warming (red) or cooling (blue) with rather

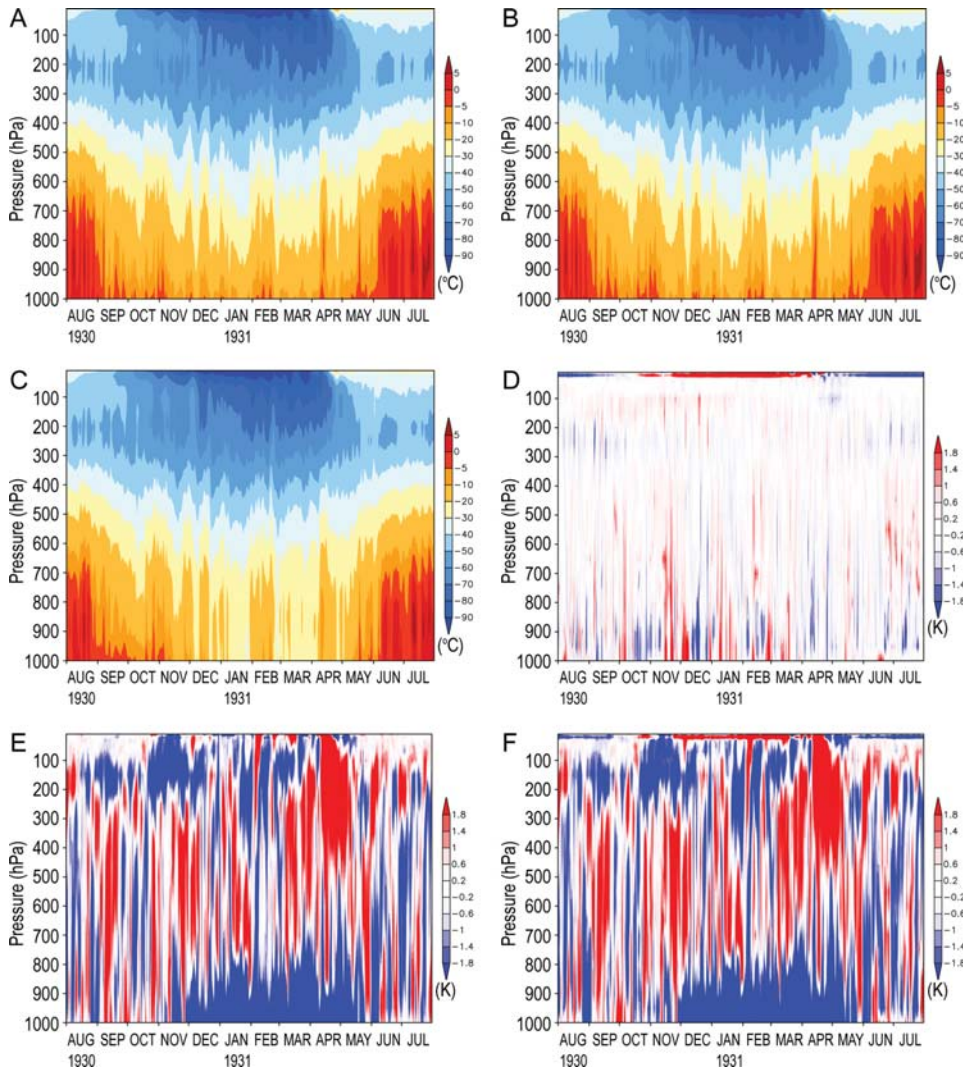


Fig. 6. Time height cross sections of daily mean air temperature from (A) 20CRv2; (B) HIRHAM5; (C) 20CRv2c, and corresponding difference plots (D) 20CRv2-HIRHAM5; (E) 20CRv2c-20CRv2; (F) 20CRv2c-HIRHAM5 for Calm Bay from August 1930 to July 1931. Model data were interpolated to the same 24 pressure levels as in the files provided by ESRL PSD from NOAA.

irregular dependency on month and altitude. Particularly in March and April 1931, 20CRv2c is clearly colder below 900 hPa but warmer between 200–300 hPa. This gives evidence of changed baroclinicity and could mean that the coupling between PBL and free troposphere is considerably biased in 20CR. The latter is investigated more detailed in the next section, which presents a more convincing comparison to observed temperature profiles.

Multi-year monthly comparison to observed temperature profiles. — Following Table 1, Fig. 7 illustrates observed (PANGAEA), reanalyzed (20CRv2, 20CRv2c), and simulated (HIRHAM5) multi-year monthly vertical profiles of temperature. Only those reanalyzed and simulated temperature profiles that have also been measured were part of the comparison. First, the daily reanalysis and model data were adapted by inserting the same missing values as included in the derived daily PANGAEA data set. This takes into account both the unavailable and the vertically interrupted balloon and radiosonde ascents as explained in the section *Vertical profiles of air temperature*. The multi-year mean calculation was performed as last step.

Below 700 hPa, Fig. 7 shows generally larger temperature differences during polar night then during polar day. This basically agrees with results shown in Fig. 2C and Fig. 5C. The improvements of HIRHAM5 compared to its forcing data (20CRv2) are rather small but visible, especially in February, November, and December. In contrast, 20CRv2c is able to correct the temperature profile not only qualitatively but shows also the smallest temperature bias in lower levels, except for July and September.

Above 400 hPa, Fig. 7 shows generally much larger differences to the observed temperatures, independent of whether compared to 20CR or HIRHAM5. In fact, this systematic upper-level cold bias can have values up to 30 K (in April), while the predominant lower-level (below 700 hPa) warm bias is limited maximally to 12 K (in March). Besides, there is often a qualitative mismatch between the reanalyzed/simulated and observed temperature profiles above 400 hPa, especially from March to June and in December. HIRHAM5 and its forcing data set (20CRv2) produce an almost identical vertical temperature structure in upper levels. Interestingly, the newer version of 20CR, *i.e.* 20CRv2c, behaves only slightly different here, except for May.

In general, the two reanalyzed and the simulated temperature profiles agree fairly well with the observations between 400–700 hPa. While all temperature profiles are qualitatively identical in this altitude range, 20CRv2c occasionally shows a slightly larger bias to the observations, especially in September.

Discussion

Reanalyzed and simulated performance of surface meteorological conditions. — Compared to the older version of 20CR, *i.e.* 20CRv2, the improvement of c_{ice} and d_{ice} , explicitly illustrated in Figs. 8D to 8E and Fig. 9, modify the absolute value of SLP but cannot change the near-surface circulation in the newer reanalysis version (20CRv2c). The latter is confirmed by daily values of SLP (Fig. 2A) and particularly by its monthly means (Fig. 8A).

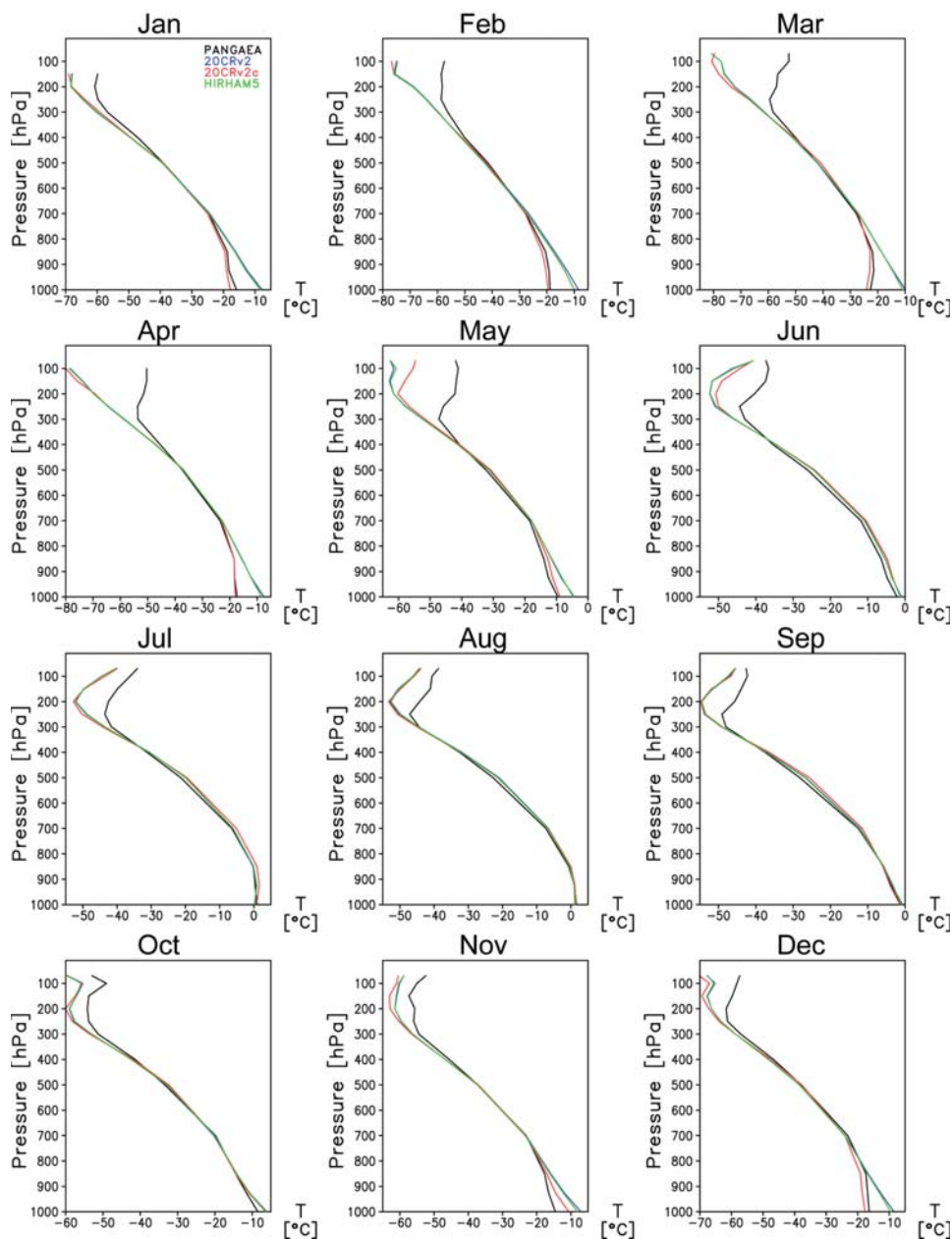


Fig. 7. Multi-year (10 September 1934 – 31 December 1940) monthly temperature profiles for Calm Bay from PANGAEA observations (black), the newer (20CRv2c in red) and older (20CRv2 in blue) version of 20CR, and the HIRHAM5 model in green. The calculation of monthly means always exploited the missing values from the daily PANGAEA data to take into account unavailable or vertically interrupted temperature profiles also in the reanalysis and model data. Temperatures are shown on the 16 standard pressure levels used by PANGAEA.

The inability of 20CR and HIRHAM5 to resolve the local topography partly explains the differences to the observed wind field. The relatively small correlations to the observed V_{10m} (Figs. 3D to 3F) mean a biased occurrence of storm events in 20CR and HIRHAM5 data. Their generally too narrow 2σ ranges in Fig. 4B argue also for a significant underestimation of storm variability. Furthermore, the systematically and significantly larger V_{10m} in the observational data from October 1930 to January 1931 (Fig. 5B) give evidence that gale force, *i.e.* cyclone strength, is underestimated by 20CR and HIRHAM5. Brönnimann *et al.* (2013) already found a rather poor agreement between 20CR and *in-situ* upper-air wind measurements performed in 1912 and 1913 on Spitsbergen. For the North Atlantic, Krueger *et al.* (2013) demonstrated that 20CR has problems in reproducing the variability and long-term trends in storminess.

Although all presented analyzes refer to a single land point (Calm Bay), it is sometimes advisable to imagine the grid cell in the model world. Thus, the actual values of c_{ice} , d_{ice} , and SST crucially determine the different T_{2m} in 20CR and HIRHAM5. An increased value of c_{ice} enhances the isolating effect and reduces or even prevents the vertical exchange of heat and moisture from the Arctic Ocean into the PBL (*e.g.*, Vihma 2014). On the other hand, a decreased c_{ice} increases or even enables mentioned interchange. Otherwise, the thinner the sea ice the larger is the conductive heat flux through the ice.

Consequently, HIRHAM5 produces smaller T_{2m} than 20CRv2, because the assumed 2 m thick sea ice is one order of magnitude thicker than the 20 cm in the older reanalysis version (Fig. 9B). The changes in T_{2m} related to the transition from 20CRv2 to 20CRv2c are consistent with the changes in SLP , because colder temperatures in 20CRv2c coincide with higher $SLPs$ (*e.g.* Figs. 4A and 4C). Despite the use of different time periods and areas, Przybylak *et al.* (2013) also found a positive bias for SLP from 20CR in historical times. The systematic cold bias in 20CRv2 concluded by Przybylak *et al.* (2016) from their fig. 11a is not reconfirmed by our results (Fig. 2C). HIRHAM5 can slightly reduce the positive temperature bias relative to its forcing data set (20CRv2) during polar night. This cooling is even more obvious in Fig. 8B, however, T_{2m} is captured most realistically by 20CRv2c.

Colder air can absorb less water vapor. Particularly Figs. 2C to 2D and Figure 4C to 4D suggest that the improved performance of 20CRv2c in reproducing SH_{2m} is rather limited to reduced biases in T_{2m} . Figure 3J demonstrates that the data points of Fig. 3J can be better represented by different regression lines for the cold and warm periods even though their correlation to observations is almost the same due to a large correlation in general. Otherwise, 20CRv2 shows the worst performance for both T_{2m} and SH_{2m} highlighting the added value of higher horizontal resolution and obviously more realistic sub-grid scale parameterizations in the HIRHAM5 model. Nevertheless, Fig. 5D shows that the air in 2 m altitude is systematically too moist, independently of whatever reanalysis or model.

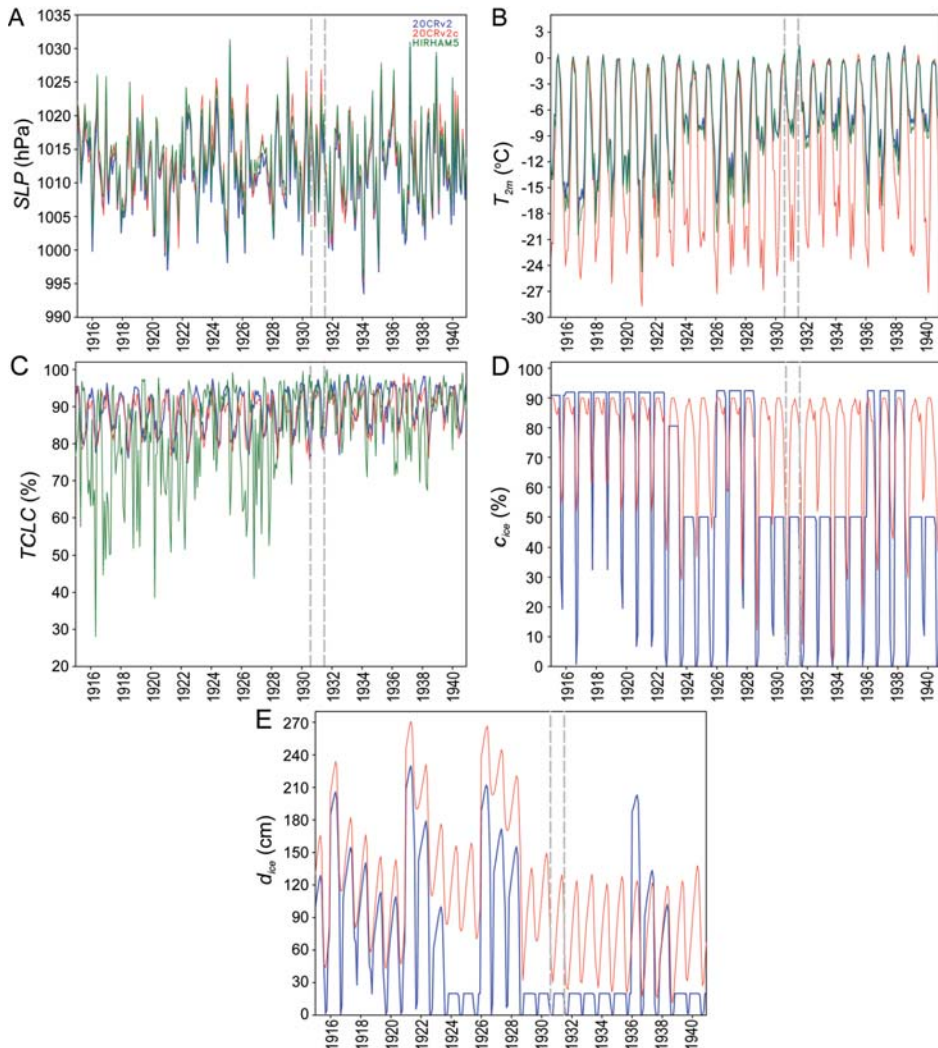


Fig. 8. Monthly means of the (A) sea level pressure; (B) 2 m air temperature; (C) total cloud cover; (D) sea ice concentration; (E) sea ice thickness produced by the newer 20CRv2c reanalysis (red), its older version 20CRv2 (blue), and the HIRHAM5 model (green) at Calm Bay from January 1915 to December 1940. Due to the lower boundary forcing, the sea ice concentration of 20CRv2 is identical in HIRHAM5, but the model assumes 2 m thick ice for the part of the grid cell that is covered with sea ice. The vertical dashed lines mark the time period, where near-surface observations are available (from August 1930 to July 1931).

Visual observations of *TCLC* need to be taken with caution due to rather subjective assignments with possible “measurement uncertainties” of at least $\pm 10\%$ and particularly difficult cloud detection capabilities during polar night. Furthermore, their comparison to reanalyzed or simulated values is considerably

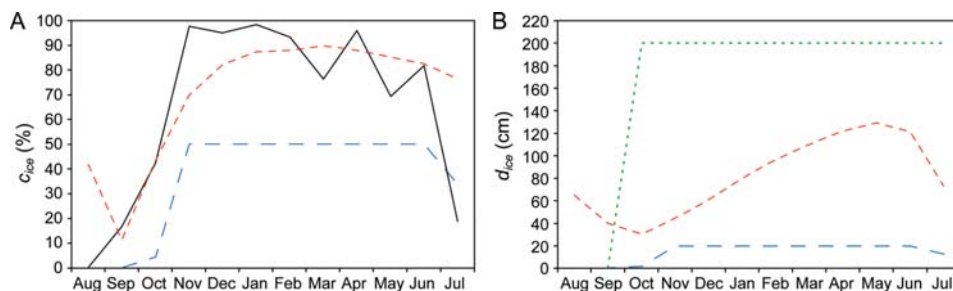


Fig. 9. Monthly means of the (A) sea ice concentration; (B) sea ice thickness from observations by Walsh *et al.* (2017; solid), the newer version of 20CR (20CRv2c, short-dashed), its older version (20CRv2, long-dashed), and the model (HIRHAM5, dotted) for Calm Bay from August 1930 to July 1931. Due to the lower boundary forcing, the sea ice concentration of 20CRv2 is identical in HIRHAM5, but the model assumes 2 m thick ice for the part of the grid cell that is covered with sea ice.

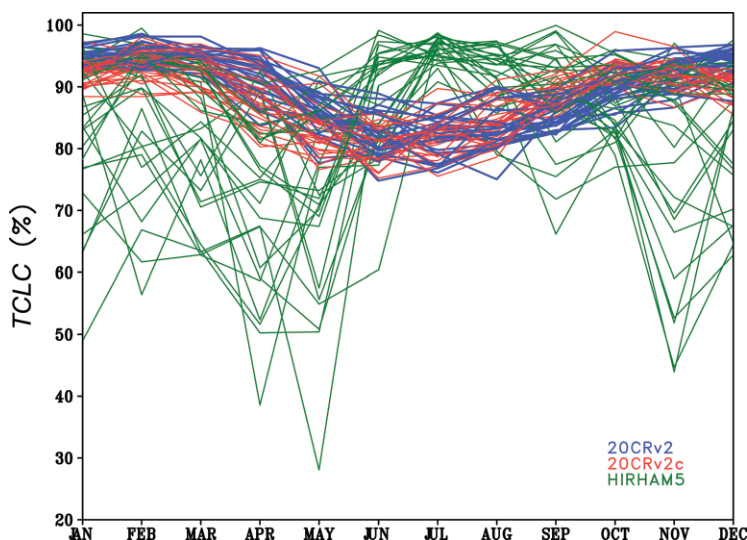


Fig. 10. Reanalyzed 20CRv2 and 20CRv2c as well as simulated HIRHAM5 annual cycle of monthly total cloud cover for each year of the entire simulation period 1915–1940.

uncertain, because the observed *TCLC* refers to the visible sky rather than to a grid cell of certain area, *e.g.* about 25×25 km² for HIRHAM5.

A high correlation on a daily time scale between observed and simulated *TCLC* cannot be expected due to various reasons. For instance, the uncertain lower BCs (*e.g.*, c_{ice} from 20CRv2) and limited forcing data for the nudging (20CRv2 is *e.g.* too moist) cannot be compensated or corrected by the RCM, even so the HIRHAM5 simulation is partly improved. In relation to the bias, Fig. 5E indicates that the performance in reproducing *TCLC* is generally poor.

However, Fig. 4E suggests a partly better month-to-month variability of *TCLC* in HIRHAM5, even if the absolute values disagree with visual observations.

Although not apparent in case of the overwintering 1930–1931 (Fig. 4E), the Fig. 10 for the entire simulation period (1915–1940) demonstrates that HIRHAM5 produces mainly a realistic annual cycle of *TCLC* with more Arctic clouds during summer and autumn but less clouds during winter. In contrast, 20CRv2 and 20CRv2c show generally a reverse annual cycle of *TCLC*. Note that the differences in *TCLC* between the model and the two versions of 20CR are much larger for other years of the simulation period, but visual cloud observations for Calm Bay are only available from 1930 to 1931, when *TCLC* differs rather slightly.

More importantly, only the model produces a positive trend in *TCLC* (Fig. 8C) with a remarkably sharp transition to enhanced cloudiness around 1928. This increase in *TCLC* is basically in agreement to increased open water areas during the ETCW that have possibly favored increased upward latent heat fluxes and enhanced cloud formation. On the other hand, the recently published study of Wegmann *et al.* (2016) indicates that both the Atlantic and the Pacific sectors were convergence zones for southerly air masses. This might be related to an intensified horizontal transport of warmer and particularly moister air, which could have favored enhanced cloud formation as well.

Reanalyzed and simulated performance of vertical temperature structure.

— The differences in the vertical temperature structure between HIRHAM5 and 20CRv2 (Fig. 6D) mainly arise from different horizontal resolutions and the use of different sub-grid scale parameterizations, *i.e.* radiation transfer, cloud formation, precipitation processes, turbulence, *etc.* The latter result, *e.g.* in different cloud amounts (Figs. 2E and 8C), which in turn lead to changes in surface cloud radiative forcing and surface sensible and latent heat fluxes as shown *e.g.* by Klaus *et al.* (2016). Above 700 hPa, HIRHAM5 and its forcing data set (20CRv2) show almost the same temperatures, because the nudging strength increases linearly with altitude.

For altitudes below 700 hPa, Fig. 7 demonstrates that the improved lower boundary forcing and the more reasonable values of d_{ice} (shown in Fig. 8E) lead mainly to more realistic temperatures in 20CRv2c. Nevertheless, this surface control is negligible above 400 hPa, manifesting in frequent qualitative mismatches, large temperature differences to the observations and a systematic upper-level cold bias independent of the comparative data. This indicates an incorrect coupling between PBL and free troposphere. For the generation of 20CR, it is therefore insufficient to assimilate only the measured surface pressure. An additional assimilation of available vertical profiles (temperature, humidity, wind speed), *e.g.* from PANGAEA or other historical radio sounding data, would further improve 20CRv2c. That the consideration of radiosonde data improves model-based circulation patterns was demonstrated by Inoue *et al.* (2013).

Reliability of lower boundary conditions (BCs). — The model shows predominantly an improved performance compared to its forcing data set (20CRv2). Due to the use of higher horizontal resolution and apparently more realistic sub-grid scale parameterizations, HIRHAM5 can reduce several biases relative to the observations. Nevertheless, the key limitations of 20CRv2, *e.g.* unrealistic sea ice, which are imprinted by the initialization, lateral boundary forcing, and altitude-dependent nudging, cannot be corrected. This reconfirms the crucial necessity of high-quality BCs obtained from a global circulation model to realistically force a RCM (Rummukainen 2010). Moore and Babij (2016) recently discussed that the correct representation of the SST and sea ice fields plays an important role for the performance of the ERA-20C reanalysis.

Here, at least for low-level temperature and near-surface specific humidity, 20CRv2c is in the best agreement to the surface-based observations. However, the improved reanalysis of SH_{2m} is rather limited to smaller biases in T_{2m} , which appear only during polar night. The application of more realistic sea ice concentrations and SSTs as lower BCs to generate 20CRv2c are comparable to a model tuning, while the physical processes in the underlying NCEP model are as defective as in case of the older version of 20CR. These fundamental errors affect the entire reanalysis system and manifest in the largest SLP bias and an almost unchanged systematic underestimation of V_{10m} compared to 20CRv2, HIRHAM5, and the observations.

In principle, the reliability of the sea ice data used to generate 20CRv2c remains a great unknown. Figure 9A explicitly illustrates a comparison between the prescribed sea ice concentration from COBE-SST2 and solely observational data from Walsh *et al.* (2016) revealing significant differences for the analyzed one-year overwintering. Nowadays, Arctic sea ice generally reaches its maximum extent in March and its minimum extent in September, *e.g.* Stroeve *et al.* (2008), which is basically captured by 20CRv2c (Fig. 9A). Nevertheless, the question arises whether this behavior is automatically applicable to historic times prior to the satellite era. Though, the observations from Walsh *et al.* (2016) rather indicate the minimum sea ice concentration in August and maximum in January. A major shortcoming in 20CR and HIRHAM5 is the omission of leads, which have a crucial impact on temperature in the PBL, especially during polar night (*e.g.*, Lüpkes *et al.* 2008).

Conclusions

The supporting pillar of this article is undoubtedly the application of the unique and independent observational data collected during the overwintering 1930–1931 at Calm Bay as reference for the evaluation of the older (20CRv2) and newer (20CRv2c) versions of the 20th Century Reanalysis and the HIRHAM5 model. However, a general discussion on spatial representativeness of the biases is

impossible due to this kind of single-point evaluation and the limited availability of observations during the ETCW period.

The model shows predominantly an improved performance compared to its forcing data set (20CRv2) due to the use of higher horizontal resolution and apparently more realistic sub-grid scale parameterizations. HIRHAM5 can slightly reduce the positive T_{2m} and SH_{2m} bias relative to its forcing data set (20CRv2) during polar night. However, T_{2m} and SH_{2m} is captured most realistically by 20CRv2c, yet the improved reanalysis of SH_{2m} is rather limited to smaller biases in T_{2m} , which appear only during polar night. HIRHAM5 produces, for the entire simulation period 1915–1940, mainly a realistic annual cycle of $TCLC$ with more Arctic clouds during summer and autumn but less clouds during winter. In contrast, 20CRv2 and 20CRv2c show generally a reverse annual cycle of $TCLC$.

The reliability of the sea ice data used to generate 20CRv2c remains a great unknown. The application of more realistic c_{ice} and SSTs as lower BCs to generate 20CRv2c are comparable to a model tuning. Compared to 20CRv2, the improvement of c_{ice} and d_{ice} modify the absolute value of SLP but cannot change the near-surface circulation in 20CRv2c. The systematically and significantly larger V_{10m} in the observational data give evidence that storm variability, *i.e.* cyclone strength, is underestimated by 20CR and HIRHAM5.

The different lower BCs, *i.e.* c_{ice} and SST, and the more realistic values of d_{ice} change the vertical stratification and baroclinicity in the transition from 20CRv2 to 20CRv2c, although the SLP is very similar in both versions of this reanalysis product. Compared to observed vertical temperature profiles from PANGAEA, the systematic upper-level above 400 hPa cold bias remains almost unchanged indicating an incorrect coupling between PBL and free troposphere.

Our results clearly suggest that beside the more realistic sea ice, an additional assimilation of available vertical profiles of temperature, humidity, and wind speed would be wise to further improve the next generation of 20CR accompanied by more reliable information about the past climate. This is a prerequisite for correctly identifying the causes of ETCW in the Arctic. Consequently, the identification and digitization of additional historical observations (Cullather *et al.* 2016), especially for the ETCW period, is critically needed for improving the 20CR reanalysis system and for model evaluation in general.

Acknowledgments. — This study was financially supported by the Helmholtz Climate Initiative REKLIM. Collecting, digitizing and analyzes of observational data were financially supported by the National Science Centre, Poland (Decision No. DEC-2012/07/B/ST10/04002). A work of Rajmund Przybylak was supported by the grant No. 2015/19/B/ST10/02933 funded by the National Science Centre, Poland. Klaus Dethloff and Annette Rinke acknowledge the support by the project *Quantifying rapid climate change in the Arctic: Regional feedbacks and large-scale impacts (QUARCCS)* funded by the German Federal Ministry of Education and Research (BMBF). We thank for the

20th Century Reanalysis v2 and v2c data, available at <http://www.esrl.noaa.gov/psd/>, provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA. Also, we are grateful for the provision of the PANGAEA archive, available at <https://doi.pangaea.de/10.1594/PANGAEA.823609>. The HIRHAM5 model simulation was conducted on the high performance computer “blizzard” of the German Climate Computing Center (DKRZ) and model data can be provided in case of request.

References

- ALEKSEEV G.V., DANILOV A.I., KATTSOV V.M., KUZMINA S.I. and IVANOV N.E. 2009. Changes in the climate and sea ice of the Northern Hemisphere in the 20th and 21st centuries from data of observations and modeling. *Izvestiya Atmospheric and Oceanic Physics* 45: 675–686.
- BENGTSOON L., SEMENOV V. and JOHANNESSEN O. 2004. The Early Twentieth-Century Warming in the Arctic – A Possible Mechanism. *Journal of Climate* 17: 4045–4057.
- BERRISFORD P., DEE D., FIELDING K., FUENTES M., KÄLLBERG P., KOBAYASHI S. and UPPALA S. 2009. *The ERA-Interim Archive*. ERA report series, European Center for Medium-Range Weather Forecasts (ECMWF): Shinfield Park, Reading UK: 20 pp.
- BRÖNNIMANN S., ANNIS J., DANN W., EWEN T., GRANT A.N., GRIESSER T., KRÄHENMANN S., MOHR C., SCHERER M. and VOGLER C.A. 2006. Guide for digitising manuscript climate data. *Climate of the Past* 2: 137–144.
- BRÖNNIMANN S., WEGMANN M., WARTENBURGER R. and STICKLER A. 2013. Arctic Winds in the “Twentieth Century Reanalysis”. In: S. Brönnimann and O. Martius (eds) *Weather extremes during the past 140 years*. *Geographica Bernensia* G89: 59–67.
- CAPPELEN J. 2009. *DMI monthly climate data collection 1768–2008, Denmark, The Faroe Islands and Greenland*. Technical Report 09-05. Danish Meteorological Institute. Copenhagen: 53 pp.
- COMPO G.P. et al. 2015. *The International Surface Pressure Databank version 3*. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. Available online: <http://dx.doi.org/10.5065/D6D50K29> (accessed on 21 Dec 2016).
- COMPO G.P., WHITAKER J.S., SARDESHMUKH P.D., MATSUI N., ALLAN R.J., YIN X., GLEASON B.E., VOSE R.S., RUTLEDGE G., BESSEMOULIN P., BRÖNNIMANN S., BRUNET M., CROUTHAMEL R.I., GRANT A.N., GROISMAN P.Y., JONES P.D., KRUK M.C., KRUGER A.C., MARSHALL G.J., MAUGERI M., MOK H.Y., NORDLI Ø., ROSS T.F., TRIGO R.M., WANG X.L., WOODRUFF S.D. and WORLEY S.J. 2011. The Twentieth Century Reanalysis Project. *Quarterly Journal of the Royal Meteorological Society* 137: 1–28.
- CULLATHER R., HAMILL T., BROMWICH D., WU X. and TAYLOR P. 2016. *Systematic Improvements of Reanalyses in the Arctic (SIRTA) White Paper*. Developed for the Interagency Arctic Research Policy Committee (IARPC). Available online: <http://www.iarpccollaborations.org/uploads/> (accessed on 21 Dec 2016).
- DAVIES H.C. 1976. A lateral boundary formulation for multi-level prediction models. *Quarterly Journal of the Royal Meteorological Society* 102: 405–418.
- DAVIES H. and TURNER R.E. 1977. Updating prediction models by dynamical relaxation: An examination of the technique. *Quarterly Journal of the Royal Meteorological Society* 103: 225–245.
- GIESE B.S., SEIDEL H.F., COMPO G.P. and SARDESHMUKH P.D. 2016. An ensemble of ocean reanalyses for 1815–2013 with sparse observational input. *Journal Geophysical Research Oceans* 121: 6891–6910.

- HEGERL G.C., CROWLEY T.J., BAUM S.K., KIM K.-Y. and HYDE W.T. 2003. Detection of volcanic, solar and greenhouse gas signals in paleoreconstructions of Northern Hemispheric temperature. *Geophysical Research Letters* 30: 1242.
- HIRAHARA S., ISHII M. and FUKUDA Y. 2014. Centennial-scale sea surface temperature analysis and its uncertainty. *Journal of Climate* 27: 57–75.
- INOUE J., ENOMOTO T. and HORI M.E. 2013. The impact of radiosonde data over the ice-free Arctic Ocean on the atmospheric circulation in the Northern Hemisphere. *Geophysical Research Letters* 40: 864–869.
- JAISER R., DETHLOFF K., HANDORF D., RINKE A. and COHEN J. 2012. Impact of sea ice cover changes on the Northern Hemisphere atmospheric winter circulation. *Tellus A* 64: 11595.
- JAISER R., NAKAMURA T., HANDORF D., DETHLOFF K., UKITA J. and YAMAZAKI K. 2016. Atmospheric winter response to Arctic sea ice changes in reanalysis data and model simulations. *Journal of Geophysical Research: Atmospheres* 121: 7564–7577.
- JOHANNESSEN O.M., BENGTTSSON L., MILES M.W., KUZMINA S.I., SEMENOV V.A., ALEKSEEV G.V., NAGURNYI A.P., ZAKHAROV V.F., BOBYLEV L.P., PETTERSSSEN L.H., HASSELMANN K. and CATTLE H.P. 2004. Arctic climate change: observed and modelled temperature and sea-ice variability. *Tellus A* 56: 328–341.
- KLAUS D., DETHLOFF K., DORN W., RINKE A. and WU D.L. 2016. New insight of Arctic cloud parameterization from regional climate model simulations, satellite-based, and drifting station data. *Geophysical Research Letters* 43: 5450–5459.
- KÖLTZOW M. 2007. The effect of a new snow and sea ice albedo scheme on regional climate model simulations. *Journal of Geophysical Research* 112: D07110.
- KRUEGER O., SCHENK F., FESER F. and WEISSE R. 2013. Inconsistencies between long-term trends in storminess derived from the 20CR reanalysis and observations. *Journal of Climate* 26: 868–874.
- LÜPKES C., VIHMA T., BIRNBAUM G. and WACKER U. 2008. Influence of leads in sea ice on the temperature of the atmospheric boundary layer during polar night. *Geophysical Research Letters* 35: L03805.
- LVOV B.C. (ed.) 1933. *Meteorological Observations of the Polar Station at Franz Josef Land, Calm Bay, Wintering of 1930-1931*. Central administration of the Hydrometeorological Service of the U. S. S. R., Central Geophysical Observatory, Section of Polar and High Altitude Observations of the Institute of Climatology, Leningrad: 31 pp.
- MEEHL G.A., WASHINGTON W.M., WIGLEY T.M.L., ARBLASTER J.M. and DAI A. 2003. Solar and greenhouse gas forcing and climate response in the twentieth century. *Journal of Climate* 16: 426–444.
- MOORE G.W.K. and BABIJ M. 2016. Iceland's Great Frost Winter of 1917/1918 and its representation in reanalyses of the twentieth century. *Quarterly Journal of the Royal Meteorological Society* 143: 508–520.
- NCAR – National Center for Atmospheric Research Stuff (eds) Last modified 30 March 2017. *The Climate Data Guide: NOAA 20th-Century Reanalysis, Version 2 and 2c*. Available online: <https://climatedataguide.ucar.edu/climate-data/noaa-20th-century-reanalysis-version-2-and-2c> (accessed on 21 Dec 2016).
- NAKAMURA T., YAMAZAKI K., IWAMOTO K., HONDA M., MIYOSHI Y., OGAWA, Y. and UKITA J. 2015. A negative phase shift of the winter AO/NAO due to the recent Arctic sea-ice reduction in late autumn. *Journal of Geophysical Research: Atmospheres* 120: 3209–3227.
- NOZAWA T., NAGASHIMA T., SHIOGAMA H. and CROOKS S.A. 2005. Detecting natural influence on surface air temperature change in the early twentieth century. *Geophysical Research Letters* 32: L20719.

- POLI P., HERSBACH H., TAN D.G.H., DEE D.P., THEPAUT J.-J., SIMMONS A., PEUBEY C., LA-LOYAUX P., KOMORI T., BERRISFORD P., DRAGANI R., TRÉMOLET Y., HOLM E., BONAVITA M., ISAKSEN L. and FISHER M. 2013. *The data assimilation system and initial performance evaluation of the ECMWF pilot reanalysis of the 20th-century assimilating surface observations only (ERA-20C)*. ERA report series 14, European Center for Medium-Range Weather Forecasts (ECMWF). Shinfield Park, Reading: 59 pp.
- POLI P., HERSBACH H., DEE D.P. and BERRISFORD P. 2016. ERA-20C: An Atmospheric Reanalysis of the Twentieth Century. *Journal of Climate* 29: 4083–4097.
- PRZYBYLAK R. 2002. *Variability of Air Temperature and Atmospheric Precipitation During a Period of Instrumental Observation in the Arctic*. Kluwer Academic Publishers, Boston-Dordrecht-London: 330 pp.
- PRZYBYLAK R. 2016. *The Climate of the Arctic*. 2nd edition. Springer: Cham–Heidelberg–New York–Dordrecht–London: 287 pp.
- PRZYBYLAK R., WYSZYŃSKI P., NORDLI Ø. and STRZYŻEWSKI T. 2016. Air temperature changes in Svalbard and the surrounding seas from 1865 to 1920. *International Journal of Climatology* 36: 2899–2916.
- PRZYBYLAK R., WYSZYŃSKI P., VÍZI Z. and JANKOWSKA J. 2013. Atmospheric pressure changes in the Arctic from 1801 to 1920. *International Journal of Climatology* 33: 1730–1760.
- RAMELLA PRALUNGO L., HAIMBERGER L., STICKLER A. and BRÖNNMANN S. 2014. A global radiosonde and tracked balloon archive on 16 pressure levels (GRASP) back to 1905 – Part 1: Merging and interpolation to 00:00 and 12:00 GMT. *Earth System Science Data* 6: 185–200.
- RAYNER N.A., PARKER D.E., HORTON E.B., FOLLAND C.K., ALEXANDER L.V., ROWELL D.P., KENT E.C. and KAPLAN A. 2003. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research – Atmosphere* 108: 4407.
- ROECKNER E., BÄUML G., BONAVENTURA L., BROKOPF R., ESCH M., GIORGETTA M., HAGEMANN S., KIRCHNER I., KORNBLUEH L., MANZINI E., RHODIN A., SCHLESE U., SCHULZWEIDA U. and TOMPKINS A. 2003. *The Atmospheric General Circulation Model ECHAM5-Part I: Model Description*; Technical Report 349. Max Planck Institute (MPI) for Meteorology, Hamburg: 140 pp.
- RÓZDŻYŃSKI K. 1999. *Meteorological metrology. Vol. 1*. Institute of Meteorology and Water Management, Warsaw (in Polish): 340 pp.
- RUMMUKAINEN M. 2010. State-of-the-art with regional climate models. *WIREs Climate Change* 1: 82–96.
- SCHERHAG R. 1937. Die Erwärmung der Arktis. *ICES Journal of Marine Science* 12: 263–276.
- SCOTT P.A., JONES G.S. and MITCHELL J.F.B. 2003. Do models underestimate the solar contribution to recent climate change? *Journal of Climate* 16: 4079–4093.
- SEMENOV V.A. and LATIF M. 2012. The early twentieth century warming and winter Arctic sea ice. *The Cryosphere* 6: 1231–1237.
- SHIOGAMA H., NAGASHIMA T., YOKOHATA T., CROOKS S.A. and NOZAWA T. 2006. Influence of volcanic activity and changes in solar irradiance on surface air temperatures in the early twentieth century. *Geophysical Research Letters* 33: L09702.
- STROEVE J., SERREZE M., DROBOT S., GEARHEARD S., HOLLAND M., MASLANIK J., MEIER W. and SCAMBOS, T. 2008. Arctic Sea Ice Extent Plummets in 2007. *EOS Transactions, American Geophysical Union* 89: 13–14.
- SUO L., OTTERA O.H., BENTSEN M., GAO Y. and JOHANNESSEN O.M. 2013. External forcing of the early 20th century Arctic warming. *Tellus A* 65: 20578.

- UNDÉN P., RONTU L., JÄRVINEN H., LYNCH P., CALVO J., CATS G., CUXART J., EEROLA K., FORTILIUS C., GARCIA-MOYA J.A., JONES C., LENDERLINK G., McDONALD A., McGRATH R., NAVASCUES B., NIELSEN N.W., ØDEGAARD V., RODRIGUEZ E., RUMMUKAINEN M., RÕÖM R., SATTLER K., SASS B.H., SAVIJÄRVI H., SCHREUR B.W., SIGG R., THE H. and TIJM A. 2002. HIRLAM-5 Scientific Documentation. In: *HIRLAM-5 Project*. Swedish Meteorological and Hydrological Institute (SMHI), Norrköping: 144 pp.
- VIHMA T. 2014. Effects of Arctic Sea Ice Decline on Weather and Climate: A Review. *Surveys in Geophysics* 35: 1175–1214.
- WALSH J.E., FETTERER F., STEWART J.S. and CHAPMAN W.L. 2017. A database for depicting Arctic sea ice variations back to 1850. *Geographical Review* 107: 89–107.
- WEGMANN M., BRÖNNIMANN S. and COMPO G.P. 2017. Tropospheric circulation during the early twentieth century Arctic warming. *Climate Dynamics* 48: 2405–2418.
- WINTON M. 2000. A reformulated three-layer sea ice model. *Journal of Atmospheric and Oceanic Technology* 17: 525–531.
- WOOD K.R. and OVERLAND J.E. 2010. Early 20th century Arctic warming in retrospect. *International Journal of Climatology* 30: 1269–1279.
- ZAKHAROV V.F. 2003. Variations of sea ice area during XX century from historical data. In: L.P. Bobylev, K.A. Kondratyev and O.M. Johannessen (eds.) *Arctic Environment Variability in the Context of the Global Change*. Springer: Berlin–Haidelberg: 107–236.
- ZUBOV N.N. 1948. *Arctic ice and the warming of the Arctic, being chapters VI and VII of “In the Center of the Arctic”: an outline of the history of Arctic exploration and of the physical geography of the central Arctic*. Northern Sea Route Directorate Press, Moscow–Leningrad: 72 pp.

Received 7 September 2017

Accepted 14 March 2018